## CONTENTS

Acknowledgement	d
Abstract in Thai	
Abstract in English	g
List of Tables	1
List of Figures	n
List of Abbreviations	0
List of Symbols	р
Statement of Originality in Thai	r
Statement of Originality in English	S
Chapter 1 Introduction	1
Chapter 1 Introduction	1
1.1 Historical Background and Motivation	1
1.2 Literature Review	5
1.2.1 General learning model	5
1.2.2 Support vector machine and the best fitting hyperplane classifiers	8
1.2.3 Data description using spherical shapes	10
1.2.4 Minimum volume covering ellipsoid	11
1.2.5 Ellipsoids in kernel feature space	14
1.3 Purposes of the Study	15
1.4 Research Scope and Method	15
1.5 Education/Application Advantages	15
1.6 Research Methodologies	16
1.7 Organization of Dissertation	16
Chapter 2 Relevant Background	18

2.1 Notation	18
2.2 Representations of an Ellipsoid	19
2.3 Minimum Volume Covering Ellipsoid	19
2.4 Support Vector based Classifiers	23
2.4.1 Support vector machine	23
2.4.2 Twin support vector machine	25
2.5 Support Vector based Classifiers with Hypersurface	28
2.5.1 Support vector data description	28
2.5.2 Twin-hypersphere support vector machine	31
Chapter 3 Support Vector Classifiers based on Hyperellipsoids	35
3.1 Ellipsoidal Support Vector Data Description	35
3.2 Twin Hyper-ellipsoidal Support Vector Machine	
3.2.1 Connection to MVCE with negative examples	43
3.3 Ellipsoids with Kernel Methods	44
3.3.1 Existing literature on kernelizing MVCE	46
3.3.2 Constructing MVCE in the feature space via empirical feature	
mapping	48
Chapter 4 Experimental Results and Discussion	51
4.1 One-class Classification with Ellipsoidal Support Vector Data Description	51
4.1.1 Standard datasets	51
4.1.2 Effect of hyperparameters	56
4.2 Two-class Classification with Twin Hyper-ellipsoidal Support Vector	
machine rights reserved	57
4.2.1 Artificial datasets	58
4.2.2 Standard datasets	62
4.2.3 Private datasets	67
4.2.4 TESVM with instance reduction	69
Chapter 5 Conclusion and Future Works	76
References	78

List of Publications	
Appendix A Relevant Proofs	
Appendix B Additional Experimental Results	
B.1 Supplementary Results for neSVDD	90
B.2 Accuracy Ranking Summary for TESVM	91

95

Curriculum Vitae



### LIST OF TABLES

Table 4.1 Standard datasets for one-class classification	52
Table 4.2 The comparison of the area under ROC for one-class classification	54
Table 4.3 Artificial datasets for binary classification	58
Table 4.4 Binary classification accuracy on artificial datasets	60
Table 4.5 Standard datasets for binary classification	63
Table 4.6 Classification accuracy (and its standard deviation) for linear THSVM	
vs. linear TESVM under two different scenarios	64
Table 4.7 Classification accuracy (and its standard deviation) for THSVM	
vs. TESVM with RBF kernel under two different scenarios	64
Table 4.8 Classification accuracy (and its standard deviation) for linear binary	
classifiers	66
Table 4.9 Classification accuracy (and its standard deviation) for classifiers	
with RBF kernel	66
Table 4.10 LCH pap smear dataset	68
Table 4.11 Classification accuracy (and its standard deviation) for classifiers	
with RBF kernel on LCH dataset	68
Table 4.12 Large datasets with at least 1000 examples	
Table 4.13 Classification accuracy and training time (in seconds) and their standard	
deviations for some selected datasets from Table 4.12	69
Table 4.14 Classification accuracy (with its standard deviation) of SVM: With vs.	
without instance reduction	72
Table 4.15 The average number of examples (and its standard deviation) of	
9 training folds in 10-fold cross validation	72
Table 4.16 Classification accuracy and training time (in seconds) and their standard	
deviations on the large datasets with instance reduction	74
Table 4.17 Classification accuracy and training time (in seconds) and their standard	
deviations on the small datasets with instance reduction	75

Table B.1 The comparison of the area under ROC for one-class classification90Table B.2 Accuracy ranking for Table 4.692Table B.3 Accuracy ranking for Table 4.792Table B.4 Accuracy ranking for Table 4.893Table B.5 Accuracy ranking for Table 4.993Table B.6 Accuracy ranking for Table 4.1694Table B.7 Accuracy ranking for Table 4.1794



### **LIST OF FIGURES**

Figure 1.1 General model of learning from data	6
Figure 2.1 Minimal ellipsoid covering a set of examples	19
Figure 2.2 Rough sketch of SVM's decision boundary	25
Figure 2.3 Rough sketch of TWSVM's decision boundary	27
Figure 2.4 Rough sketch of SVDD with negative examples	30
Figure 2.5 Rough sketch of THSVM's decision boundary	34
Figure 3.1 Rough sketch of eSVDD's descriptive boundary	36
Figure 3.2 Rough sketch of neSVDD's descriptive boundary	37
Figure 3.3 Rough sketch of TESVM's decision boundary	41
Figure 4.1 Effect of the parameter C on the boundary of eSVDD	56
Figure 4.2 Effect of the parameter C on the boundary of neSVDD	56
Figure 4.3 Contour plots of the effect of $\sigma$ parameter on neSVDD with RBF kernel	57
Figure 4.4 Toy 1 dataset with linear kernel. The test accuracy on training set is	
(a) 92.5% (b) 80.5% (c) 78.5% (d) 92.0%	59
Figure 4.5 Toy 2 dataset with linear kernel. The test accuracy on training set is	
(a) 80.0% (b) 99.2% (c) 97.2% (d) 97.6%.	61
Figure 4.6 Checkerboard dataset with RBF kernel. The test accuracy on training	
set is (a) 98.8% (b) 99.3% (c) 98.2% (d) 98.5%	62
Figure 4.7 Examples of cervical cancer cell (Courtesy of T. Chankong)	67
Figure 4.8 Instance reduction procedure	70
Figure 4.9 Banana dataset and its reduced version	71
Figure 4.10 Ho-Kleinberg's checkerboard dataset and its reduced version	71
Figure B.1 Differences of the area under the ROC curve	91

#### LIST OF ABBREVIATIONS

eSVDD	Ellipsoidal Support Vector Data Description
KKT	Karush–Kuhn–Tucker
MEB	Minimum Enclosing Ball
MVCE	Minimum Volume Covering Ellipsoid
neSVDD	Ellipsoidal Support Vector Data Description with negative
	examples
nSVDD	Support Vector Data Description with negative examples
PCA	Principal Component Analysis
QPP G	Quadratic Programming Problem
RBF	Radial Basis Function
ROC curve	Receiver Operating Characteristic curve
SDP	Semidefinite Program
SVD	Singular Value Decomposition
SVDD	Support Vector Data Description
SVM	Support Vector Machine
TESVM	Twin Hyper-ellipsoidal Support Vector Machine
THSVM	Twin-Hypersphere Support Vector Machine
TWSVM	Twin Support Vector Machine
ลขสา	<b>เธมหาวทยาลยเชย</b> งเหม
Convi	abt <sup>©</sup> by Chiang Mai University

Copyright<sup>©</sup> by Chiang Mai University All rights reserved

### LIST OF SYMBOLS

$\square$ <sup>n</sup>	<i>n</i> -dimensional Euclidean space
$S^n$	The space of symmetric $n \times n$ matrices
$S^n_+$	The space of symmetric positive semidefinite $n \times n$ matrices
$S^n_{\scriptscriptstyle ++}$	The space of symmetric positive definite $n \times n$ matrices
Н	Feature space
H <sub>E</sub>	Empirical feature space
× /8	Generalized inequality
M <sup>+</sup>	Moore–Penrose inverse of the matrix M
I	Identity matrix
e di	A vector of ones
	The set of training examples of Class <i>i</i>
n	The number of features
n <sub>c</sub>	The number of classes
т	The total number of examples in all classes
<i>m</i> <sub>i</sub>	The number of examples in Class <i>i</i>
$\{(\mathbf{x}_i, y_i)\}_{i=1}^m$	The set of <i>m</i> training examples
x adar	A testing example
x <sub>i</sub> Copyri	The <i>i</i> -th training example
$\boldsymbol{\varphi}(\mathbf{x}_i)$	The image of $\mathbf{x}_i$ in the feature space
${\mathcal{Y}}_i$	The corresponding label of $\mathbf{x}_i$
X	The column-wise matrix of all training examples
$\mathbf{X}_i$	The column-wise matrix of the training examples from Class <i>i</i>
У	The corresponding label vector of <b>X</b>
Y	The matrix whose diagonal elements are the vector $\mathbf{y}$
A	The matrix whose diagonal elements are the vector $\boldsymbol{\alpha}$

- $\alpha_i$ , **a** The *i*-th Lagrange multiplier and the Lagrange multiplier vector
- $\beta_i$ , **\beta** The *i*-th Lagrange multiplier and the Lagrange multiplier vector
- $\xi_i, \xi$  The *i*-th slack variable and the slack variable vector
- C,  $C_i$ ,  $v_i$  The hyperparameters of learning machines
- K The kernel matrix

 $\sigma$ 

 $\mathbf{\Omega} \qquad \text{A factorized matrix from } \mathbf{K} = \mathbf{\Omega}^T \mathbf{\Omega}$ 

The parameter of RBF kernel function as defined by



# ข้อความแห่งการริเริ่ม

- วิทยานิพนธ์นี้ได้นำเสนอวิธีการสร้างทรงรีที่มีปริมาตรน้อยที่สุดในปริภูมิที่กำหนดโดยเกอร์ เนลฟังก์ชันด้วยการใช้เทคนิกเอ็มพิริกอลฟีเจอร์แมป ซึ่งทำให้สามารถสร้างขอบเขตที่ใช้ สำหรับอธิบายลักษณะของข้อมูลได้ซับซ้อนมากขึ้น
- ในการสร้างทรงรีที่มีปริมาตรน้อยที่สุดเพื่อล้อมรอบข้อมูลนั้น มีการใช้ซอร์ฟมาร์จินในการ ผ่อนกลายเงื่อนไขของการสร้างทรงรี และไม่เพียงแต่ข้อมูลเป้าหมายที่สนใจจะถูกนำมาใช้ใน การกำนวณเท่านั้น แต่ข้อมูลที่ไม่เกี่ยวข้องก็ได้ถูกนำมาใช้ด้วยเช่นกัน
- 3) วิทยานิพนธ์นี้ได้นำเสนอตัวจำแนกข้อมูล 2 กลุ่มที่มีทรงรีเป็นพื้นฐาน ในชื่อ ทวินไฮเปอร์อีลิป ซอยดอลซัพพอร์ตเวกเตอร์แมชีน โดยเป็นการสร้างทรงรีที่มีปริมาตรน้อยที่สุดในการล้อมรอบ ข้อมูลแต่ละกลุ่ม โดยที่ทรงรีที่ล้อมรอบข้อมูลหนึ่งกลุ่มจะต้องอยู่ห่างจากข้อมูลอีกหนึ่งกลุ่มให้ มากที่สุดเท่าที่จะเป็นไปได้

 AIUNIVERSI

 AIUNIVERSI

 AIUNIVERSI

 Baansuraanse

 Baansuraanse

 Copyright<sup>©</sup>

 by Chiang Mai University

 AII rights reserved

r

#### STATEMENTS OF ORIGINALITY

- 1. A method to kernelize minimum volume covering ellipsoid using empirical feature map is proposed in order to allow minimum volume covering ellipsoid to form more complex description boundary.
- 2. Minimum volume covering ellipsoid is formulated with soft margins and negative examples.
- 3. An ellipsoid-based binary classifier called twin hyper-ellipsoidal support vector machine is proposed. The proposed method is to find a kernelized soft-margin minimum volume ellipsoid around one class, but also being as far as possible from the other class.

